Search for $H \rightarrow WW^* \rightarrow l\nu l\nu$
Based on Boosted Decision Trees

Hai-Jun Yang
University of Michigan

LHC Physics Signature Workshop
January 5-11, 2008
Outline

- H → WW* – a possible early discovery channel
- Brief Introduction of Boosted-Decision-Trees
- H → WW* → ℓνℓν analysis based on BDT
- ATLAS Sensitivity of H → WW* → ℓνℓν
- Summary and Outlook
Higgs Production at LHC

Gluon-gluon fusion and WW/ZZ fusion are two dominant Higgs production mechanisms.
Higgs Decay Branching Ratio and Discovery Channels

**Low mass region: $m(H) < 2 m_Z$**
- $H \rightarrow \gamma \gamma$
- $H \rightarrow b \bar{b}$
- $H \rightarrow \tau \bar{\tau}$
- $H \rightarrow ZZ^* \rightarrow 4\ell$
- $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ or $\ell\nu jj$

**$m(H) > 2 m_Z$**
- $H \rightarrow ZZ \rightarrow 4\ell$
- $qqH \rightarrow ZZ \rightarrow \ell\ell \nu\nu^*$
- $qqH \rightarrow ZZ \rightarrow \ell\ell jj^*$
- $qqH \rightarrow WW \rightarrow \ell\nu jj^*$
* for $m_H > 300$ GeV
  forward jet tag
H → WW* → lνlν

Current limit and discovery potential at LHC

Excluded cross section times Branching Ratio at 95% C.L.

CMS Phys. TDR 2006
ATLAS Physics ‘Commissioning’

- Study the new physics discovery potential with CSC (computing system commissioning) program (started from summer of 2006)
- Physics ‘TDR’ will be updated soon with ATLAS CSC note using many 10^{th} of Million fully simulated CSC MC data sets and with advanced analysis tools.
- We have developed and applied the BDT technique in diboson physics and Higgs discovery studies with the ATLAS CSC program.
Boosted Decision Trees (BDT)

- Relative new in HEP – MiniBooNE, BaBar, D0(single top discovery), ATLAS
- Advantages: robust, understand ‘powerful’ variables, ‘not a black box’, …

- Split data recursively based on input variables until a stopping criterion is reached (e.g. purity, too few events)
- Every event ends up in a “signal” or a “background” leaf
- Misclassified events will be given larger weight in the next tree (boosting)
- For a given event, if it lands on the signal leaf in one tree, it is given a score of 1, otherwise, -1. The sum of scores from all trees is the final score of the event.

B.P. Roe, H.J. Yang, et.al., physics/0408124, NIM A543 (2005) 577
H.J. Yang, B.P. Roe, et.al., physics/0610276, NIM A574 (2007) 342
H → WW* → lνlν (l = e, μ)

• Cross sections of H → WW* → lνlν (GGF & VBF) at LO (Pythia), K-factor ~ 1.9

<table>
<thead>
<tr>
<th>Higgs Mass</th>
<th>$\sigma_{GGF}$ (fb)</th>
<th>$\sigma_{VBF}$ (fb)</th>
<th>$\sigma_{total}$ (fb)</th>
<th>filter efficiency</th>
<th>Br(pp → H → WW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>140 GeV</td>
<td>328.2 (79%)</td>
<td>85.5 (21%)</td>
<td>413.2</td>
<td>0.9545</td>
<td>0.516</td>
</tr>
<tr>
<td>150 GeV</td>
<td>402.3 (79%)</td>
<td>109.8 (21%)</td>
<td>512.2</td>
<td>0.9573</td>
<td>0.704</td>
</tr>
<tr>
<td>160 GeV</td>
<td>467.0 (78%)</td>
<td>132.7 (22%)</td>
<td>600.3</td>
<td>0.9571</td>
<td>0.906</td>
</tr>
<tr>
<td>165 GeV</td>
<td>469.3 (77%)</td>
<td>135.7 (23%)</td>
<td>605.6</td>
<td>0.9579</td>
<td>0.960</td>
</tr>
<tr>
<td>170 GeV</td>
<td>448.2 (77%)</td>
<td>132.3 (23%)</td>
<td>580.4</td>
<td>0.9609</td>
<td>0.965</td>
</tr>
<tr>
<td>180 GeV</td>
<td>390.4 (76%)</td>
<td>119.3 (24%)</td>
<td>510.7</td>
<td>0.9657</td>
<td>0.933</td>
</tr>
</tbody>
</table>

H → WW signal and background simulations used
ATLAS software release v12 (for CSC note)
Full ATLAS detector simulation and reconstruction
Backgrounds

<table>
<thead>
<tr>
<th>Process</th>
<th>MC sample</th>
<th>cross-section</th>
</tr>
</thead>
<tbody>
<tr>
<td>$WW \rightarrow lvlv \ (l=e,\mu,\tau)$</td>
<td>372.5K,</td>
<td>11.72 pb</td>
</tr>
<tr>
<td>$gg2WW \rightarrow lvlv \ (l=e,\mu,\tau)$</td>
<td>209.1K,</td>
<td>0.54 pb</td>
</tr>
<tr>
<td>$ttbar \rightarrow l + X$</td>
<td>584.1K,</td>
<td>450.0 pb</td>
</tr>
<tr>
<td>$WZ \rightarrow l\ell l \ (l=e,\mu)$</td>
<td>281.4K,</td>
<td>0.7 pb</td>
</tr>
<tr>
<td>$Z \rightarrow ll \ (l=e,\mu,\tau)$</td>
<td>1.15 M,</td>
<td>4.6 nb</td>
</tr>
</tbody>
</table>

- $W/Z + \text{Jets}$ are potential background, using 1.1M fully simulated MC events (Alpgen generator), no event is selected in our final sample.
- Background estimate uncertainty $\sim 15 – 20 \%$. 
**H → WW Pre-selection**

- At least one lepton pair (ee, μμ, eμ) with $P_T > 10$ GeV, $|\eta| < 2.5$
- Missing $E_T > 15$ GeV
- $|M_{ee} - M_Z| > 10$ GeV, $|M_{\mu\mu} - M_Z| > 15$ GeV to suppress background from $Z \rightarrow ee, \mu\mu$

<table>
<thead>
<tr>
<th>Higgs Mass (GeV)</th>
<th>Eff(eeX)</th>
<th>Eff(μμX)</th>
<th>Eff(eμX)</th>
</tr>
</thead>
<tbody>
<tr>
<td>140</td>
<td>27.0%</td>
<td>53.9%</td>
<td>39.0%</td>
</tr>
<tr>
<td>150</td>
<td>29.2%</td>
<td>54.6%</td>
<td>41.1%</td>
</tr>
<tr>
<td>160</td>
<td>30.4%</td>
<td>56.3%</td>
<td>43.2%</td>
</tr>
<tr>
<td>165</td>
<td>31.0%</td>
<td>56.8%</td>
<td>43.8%</td>
</tr>
<tr>
<td>170</td>
<td>31.1%</td>
<td>55.4%</td>
<td>45.0%</td>
</tr>
<tr>
<td>180</td>
<td>29.7%</td>
<td>52.4%</td>
<td>45.8%</td>
</tr>
</tbody>
</table>

- IsEM & 0x7FF == 0 (tight electron id cuts)
- Staco-muon id
BDT Training with pre-selected events

Input physics variables to BDT program (1)

- Energy and Momentum
  - $p_T(\ell)$, $p_T(\ell, \ell)$
  - $MET$, total recoil $E_T$
  - scalar $\sum E_T(jet)$, vector $\sum E_T(\ell, MET)$

- Lepton Isolation
  - Number of tracks in $\Delta R < 0.4$ cone around $\ell$
  - Sum of track $p_T$ in $\Delta R < 0.4$ cone around $\ell$
  - Sum of jet $E_T$ in $\Delta R < 0.4$ cone around $\ell$
Input physics variables to BDT program (2)

- Event Topology
  - Number of Jets with $E_T > 30$ GeV
  - $E(\ell)/P(\ell)$
  - $A0$ (impact parameter) of $\ell$, $\Delta A0(\ell, \ell)$, $\Delta Z(\ell, \ell)$
  - $\Delta R(\ell, \ell)$, $\Delta \phi(\ell, \ell)$, $\Delta \phi(\ell, MET)$
  - $\Delta \Omega(\ell, \ell)$ - opening angle of two leptons

- Mass Information
  - Invariant mass($\ell, \ell$)
  - Transverse mass($\ell\ell$,MET)
  - Transverse mass($\ell$,MET)
Some Training Variable Distributions

ATLAS - after Precuts

No. of Tracks within a $\Delta R < 0.4$ cone
Some Training Variables

Number of Jets

ATLAS - after Precuts

Sum of Jet Et

ATLAS - after Precuts
Some Training Variables
H\rightarrow WW^* \rightarrow e\nu\mu\nu \ (165 \text{ GeV})

BDT output spectrum and selected signal & background events for 1fb$^{-1}$
After BDT Selection \((H \to \text{WW}^* \to e\nu\mu\nu)\)
S/B Ratio of H → WW* → ℓνℓν

ATLAS (1 fb⁻¹), H → ww → ℓνℓν

<table>
<thead>
<tr>
<th></th>
<th>eνeν</th>
<th>μνμν</th>
<th>eνμν</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG140</td>
<td>36.0</td>
<td>83.9</td>
<td>94.5</td>
</tr>
<tr>
<td>H140</td>
<td>12.8</td>
<td>31.2</td>
<td>37.2</td>
</tr>
<tr>
<td>BG150</td>
<td>26.6</td>
<td>85.6</td>
<td>81.0</td>
</tr>
<tr>
<td>H150</td>
<td>14.1</td>
<td>42.1</td>
<td>44.9</td>
</tr>
<tr>
<td>BG160</td>
<td>15.5</td>
<td>39.7</td>
<td>35.2</td>
</tr>
<tr>
<td>H160</td>
<td>17.5</td>
<td>40.2</td>
<td>46.8</td>
</tr>
<tr>
<td>BG165</td>
<td>22.2</td>
<td>43.5</td>
<td>36.4</td>
</tr>
<tr>
<td>H165</td>
<td>22.9</td>
<td>45.9</td>
<td>51.8</td>
</tr>
<tr>
<td>BG170</td>
<td>37.2</td>
<td>36.1</td>
<td>32.1</td>
</tr>
<tr>
<td>H170</td>
<td>28.2</td>
<td>35.5</td>
<td>42.8</td>
</tr>
<tr>
<td>BG180</td>
<td>34.6</td>
<td>36.8</td>
<td>59.4</td>
</tr>
<tr>
<td>H180</td>
<td>19.8</td>
<td>25.8</td>
<td>44.3</td>
</tr>
</tbody>
</table>
Discovery Confidence Level Calculation

- Log-likelihood ratio test-statistics by using BDT bins and 3 Higgs decay channels

\[ Q = \frac{L(s + b)}{L(b)} \]

- MC experiments are based on Poisson statistics

- $\text{CL}_b$ represents C.L. to exclude “background only” hypothesis
# Results (H→WW*→lνlν, for 1fb⁻¹)

<table>
<thead>
<tr>
<th>$M_{\text{Higgs}}$ (GeV)</th>
<th>$\text{Eff}<em>s = \text{Eff}</em>{\text{pre}} \times \text{Eff}_{\text{bdt}}$</th>
<th>$N_s$</th>
<th>$N_{\text{bg}}$</th>
<th>$N_\sigma$ (stat)</th>
<th>$N_\sigma$(10% syst)</th>
<th>$N_\sigma$(20% syst)</th>
<th>$N_{\sigma20}$ (-2lnQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>140</td>
<td>9.7%</td>
<td>81.2</td>
<td>214.4</td>
<td>5.5</td>
<td>3.1</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>150</td>
<td>9.9%</td>
<td>101.1</td>
<td>193.2</td>
<td>7.3</td>
<td>4.2</td>
<td>2.5</td>
<td>2.6</td>
</tr>
<tr>
<td>160</td>
<td>9.0%</td>
<td>104.6</td>
<td>90.4</td>
<td>11.0</td>
<td>8.0</td>
<td>5.1</td>
<td>5.3</td>
</tr>
<tr>
<td>165</td>
<td>10.3%</td>
<td>120.5</td>
<td>102.2</td>
<td>11.9</td>
<td>8.4</td>
<td>5.3</td>
<td>5.6</td>
</tr>
<tr>
<td>170</td>
<td>9.6%</td>
<td>106.5</td>
<td>105.4</td>
<td>10.4</td>
<td>7.2</td>
<td>4.5</td>
<td>4.7</td>
</tr>
<tr>
<td>180</td>
<td>9.2%</td>
<td>89.9</td>
<td>130.8</td>
<td>7.9</td>
<td>5.2</td>
<td>3.1</td>
<td>3.4</td>
</tr>
</tbody>
</table>
ATLAS Sensitivity of $H \rightarrow WW^* \rightarrow l\nu l\nu$

\[ Q = \frac{L(s + b)}{L(b)} \]

Log-likelihood Ratio with 20% syst. error

$N = N_{\text{signal}} / \sqrt{(N_{\text{bkgd}}^c + (\sigma_{\text{syst}} N_{\text{bkgd}}^d)^2)}$
Required Int. Lumi for 5σ Discovery

BDT Analysis, $H \rightarrow WW^* \rightarrow l\ell\nu\nu$ ($l=e,\mu$)

$\sigma_{syst} = 19\%, 16\%, 11\%$ for 1, 2, 10 fb$^{-1}$

ATLAS (5σ), $H \rightarrow ww \rightarrow l\ell\nu\nu$

CMS Phys. TDR 2006

$L_\text{int}$ needed for a 5σ discovery [fb$^{-1}$]

$m_H$ [GeV/c$^2$]

Statistical errors

Systematics included

NLO cross sections

$\sigma_{syst} = 19\%, 16\%, 11\%$ for 1, 2, 10 fb$^{-1}$
Cross Section Uncertainty of $H \rightarrow WW^* \rightarrow l\nu l\nu$

ATLAS (1fb$^{-1}$), $H \rightarrow \nu\nu \rightarrow l\nu l\nu$

- $e\nu\nu$
- $\mu\nu\nu$
- $e\nu\mu\nu$
- $l\nu l\nu$

Uncertainty $= \sqrt{N_{total}} / N_{signal}$

Higgs Mass (GeV)
Summary and Outlook

• $H \rightarrow WW^{*} \rightarrow ll\nu\nu$ analysis based on BDT has significant impact on early discovery potential.

• For 140-180 GeV SM Higgs $5\sigma$ discovery only needs a few fb$^{-1}$ integrated luminosity.

• Major backgrounds for $H \rightarrow WW$ searches come from $WW(50-60\%)$ and $ttbar(30-40\%).$

$\Rightarrow$ BDT is anticipated to have wide application in LHC physics analysis, especially for particle searches.
Backup Slides
H$\rightarrow$WW$^*$$\rightarrow$eevv (165 GeV)
$H \rightarrow WW^* \rightarrow \mu \nu \mu \nu$ (165 GeV)
Weak $\rightarrow$ Powerful Classifier

- The advantage of using boosted decision trees is that it combines many decision trees, "weak" classifiers, to make a powerful classifier. The performance of boosted decision trees is stable after a few hundred tree iterations.

- Boosted decision trees focus on the misclassified events which usually have high weights after hundreds of tree iterations. An individual tree has a very weak discriminating power; the weighted misclassified event rate $\text{err}_m$ is about 0.4-0.45.

BDT Training with Event Reweighting

- In the original BDT training program, all training events are set to have same weights in the beginning (the first tree). It works fine if all MC processes are produced based on their production rates.

- Our MCs are produced separately, the event weights vary from various backgrounds. e.g. 1 fb⁻¹ , wt (ww)=0.07, wt (ttbar)=0.72

- If we treat all training events with different weights equally using “standard” training algorithm, ANN/BDT tend to pay more attention to events with lower weights (high stat.) and introduce training prejudice.